**SUPPLEMENTARY METHODS, INFORMATION, AND REFERENCES**
**(McDonel *et al.*)**

## METHODS

**DNA sequence datasets.** Sequence analysis was performed on 914 bp of *rex* site sequences, their containing cosmids (R160, B0294, F42E11, and F29G6, 141,916 bp total), 16 X chromosome cosmids that do not recruit the DCC in our assay (T05A10, F31B12, F52D10, C49F5, E01G6, F17E5, ZK455, F08B12, F46C3, C35C5, T04F8, C34E11, F54F7, C23H4, C04B4, C30E1, 535,109 bp total), a set of PCR products 100 bp-5 kb in length that do not recruit the DCC (23,071 bp total), the X chromosome (17,718,851 bp), the autosomes (82,551,067), and the *C. elegans* genome (100,269,918 bp). All sequences are available from www.wormbase.org, except for sequences of the non-recruiting PCR products available from http://mcb.berkeley.edu/labs/meyer/mcdonel/suppseq.fa

**Motif selection.** Among the numerous models output by MEME[1] under the large parameter space explored, motifs A and B were chosen for subsequent functional analysis for the following reasons. First, motif A was represented in every parameter set output, and B was represented in nearly all outputs, making them by far the most frequently identified motifs in our analysis. Second, motifs A and B were consistently the highest scoring (MEME log likelihood) PWMs of all output motifs. Third, motifs A and B, especially when considered together, were strikingly over-represented in *rex* sequences when compared to background sequences (X chromosomes, autosomes, and the whole genome).

**Motif scanning.** Patser[2] was used to scan each sequence dataset listed above against PWMs for motifs A and B, using a range of raw score cutoffs from 5.5 to 8.5. Motif enrichment in all scanned datasets for both PWMs was evaluated using binomial tests against empirically determined whole genome background frequencies. A raw score cutoff of 8 for both motifs was the most effective at discriminating the datasets in all subsequent analyses.

**Motif clustering analysis and predictive model.** To evaluate the hypothesis that local clusters of motifs A and B comprise at least part of the functional units for DCC recruitment, we compared the distribution of A and B motifs, both independently and when clustered, among 5 datasets: *rex* cosmids, non-recruiting cosmids, the X chromosome, the autosomes, and the entire *C. elegans* genome. We evaluated values for the following characteristics: the summed raw score of the highest-scoring A and B motifs within each sequence; the sum of motif scores for the highest-scoring motif cluster (200-600 bp window) containing multiple A motifs, B motifs, or A and B motifs together; the length-normalized sum-of-scores for all motif A and B co-clusters (200-600 bp) scoring above a range of cutoffs (7.0-8.5); the density of either motif A or motif B scoring above a range of cutoffs (7.0-8.5); and the density of motif A and B clusters, either homotypic or containing at least one motif A and one motif B above the score cutoff range of 7.0-8.5. Only two of these characteristics were found to be significant

discriminators of *rex* cosmids from the other sequences and robust to small-sample effects: (i) the sum of motif A and motif B cluster scores, normalized for cosmid or chromosome length (600 bp windows, with each motif score ≥8.0 and length-normalized sum-of-scores ≥0.5, $P < 0.001$); (ii) density of high-scoring (each motif score ≥8.0) A and B motifs clustered in 600 bp windows ($P < 0.003$). Significance was assessed by student's T-test (two-tailed, homoscedastic).

**Distribution of predicted positive windows on X.** Non-overlapping 30 kb segments of the X chromosome sequence were evaluated as follows. For all 600 bp windows containing at least one motif A and one motif B, each with a raw Patser motif score ≥8.0, motif scores with this cutoff were summed. In this analysis, 30 kb segments with summed cluster scores ≥16 (indicating at least one high-scoring motif cluster) were called as "positive."

All 30 kb windows were grouped according to whether they resided in the strongly recruiting, weakly recruiting, or non-recruiting regions[5] of X. The boundaries of each region were inferred by merging genetic mapping data, FISH mapping, and the physical map available at www.wormbase.org. Significance of the non-uniform distribution of positive segments was assessed by Chi square analysis (see Supplementary Table 2).

**Test of predictive model.** Recruitment ability of 43 individual cosmids spanning X chromosome positions 10376300 to 12421676 (region D and part of region C) was tested to assess the predictive ability of the model parameters. This cosmid set excluded the cosmid training set.

**DNA sequences of *rex-1* fragments.**

*rex-1*•33:
GGCTGCGGGTAATTGGGCAGGGGAAAGAAGAAT

*rex-1*•60:
ACGGGAGGAGAAAGATGGAGAACATGTGGCTGCGGGTAATTGGGCAGGGGA
AAGAAGAAT

*rex-1*•89:
AAGCGCAGGGAGAACTGGTGGGAGGATGGACGGGAGGAGAAAGATGGAGA
ACATGTGGCTGCGGGTAATTGGGCAGGGGAAAGAAGAAT

*rex-1*•148:
ATTATTATACCCTGCATCAACAAGCCGCAATGCAGCAGTGCGTGCGTACAAA
AGGAGACAAGCGCAGGGAGAACTGGTGGGAGGATGGACGGGAGGAGAAAG
ATGGAGAACATGTGGCTGCGGGTAATTGGGCAGGGGAAAGAAGAAT

*rex-1*•241:
GACACAATTACAATGTTTTCAAAATTTTTCATTTGTTTTTTAATGTGTTCAGCA
GTCTTGTTTCAAAATATCAGTTTTTCAGAATTTACACGCATTATTATACCCTGC

ATCAACAAGCCGCAATGCAGCAGTGCGTGCGTACAAAAGGAGACAAGCGCA
GGGAGAACTGGTGGGAGGATGGACGGGAGGAGAAAGATGGAGAACATGTG
GCTGCGGGTAATTGGGCAGGGGAAAGAAGAAT

The previously reported X recruitment element[5] overlaps *rex-1*•241 by 108 bp, including motifs $A_1$, $A_4$, $A_2$, and $B_2$. The 793 bp fragment also contains the $A_3$ motif and six additonal B motifs.

**Worm strains**. In addition to the strains listed in Table 1, the following *C. elegans* strains were used in this study. All *yEx* extrachromosomal arrays listed below contain a pharyngeal *myo-2::GFP* transformation marker.

TY4652 *her-1(hv1y101)* V; *xol-1(y9) sdc-2(y74) unc-9(e101)* X; *yEx1041* (contains 4.5 kb *rex-1* PCR product). DCC binding to arrays with the 4.5 kb *rex-1* fragment is assessed in the absence of SDC-2.

TY4121 *her-1(e1520) sdc-3(y126)* V; *xol-1(y9)* X; *yEx754* (contains 4.5 kb *rex-1* PCR product). DCC binding to arrays with the 4.5 kb *rex-1* fragment is assessed in the absence of SDC-3.

TY4158 *yIs63* is an autosomal-linked UV integrated version of *yEx736* (contains 4.5 kb *rex-1* PCR product).

TY4654 *unc-119(ed3)* III; *yIs105* [4.5 kb *rex-1* sequence in the pER723 *unc-119*(+) vector plus *myo-2::GFP*, integrated into an autosome at 20 copies by bombardment[3]]. Strain used to assess DCC binding to low-copy *rex-1* integration sites on autosomes.

TY1158 *rol-6(e187)* II; *xol-1(y9) unc-3(e151)* X Strain used to measure suppression of *xol-1* XO lethality by *rex-1* arrays.

***rex* plasmids.** *rex* fragments larger than 115 bp in length were amplified by polymerase chain reaction (PCR) using TaqPlus Precision DNA polymerase mix (Stratagene). *rex* fragments 115 bp and smaller were generated by synthesis of complementary oligonucleotides (IDT), which were annealed in 20 mM NaCl. Each blunt-ended *rex* fragment was then ligated into the 3.5 kb neutral vector pCR-Blunt II-TOPO (Invitrogen) according to supplied instructions. All inserts were verified by DNA sequencing.

**Transgenic arrays**. Adult wild-type (N2) hermaphrodites were injected with purified cosmids (5-10 μg/ml each), PCR products (10-25 μg/ml), or plasmids (10-25 μg/ml) along with the pharyngeal GFP transformation marker *myo-2::GFP* (pPD118.33, 12-15 μg/ml) diluted in water. Independent transgenic lines were established from F1s cloned onto individual plates.

**FISH probes and antibodies.** FISH probes against transgenic arrays were generated using the Prime-A-Gene random prime labeling kit (Promega), with the $Amp^R$

portion of pBluescript as a template and incorporating fluorescent nucleotides. Members of the dosage compensation complex were detected with rabbit DPY-27 antibodies (r699)[4], rat SDC-3 antibodies (PEM-4A) raised against amino acids 1067-1340 of SDC-3 fused to GST, and rabbit SDC-2 antibodies (#3778) raised against amino acids 9-455 of SDC-2 fused to 6xHis.

**FISH, immunofluorescence staining, and microscopy of whole animals (Fig. 2 and Supplementary Figs 1d, 3 and 4).** Adult transgenic animals were dissected, fixed, and stained as described[5]. Complete intestinal cell nuclei were imaged using a Leica SP-2 AOBS confocal system (Leica Microsystems, Germany) with a 63x 1.4 NA lens. Stacks of 7-14 Z-sections spaced 0.4-0.5 μm apart were captured with an xy voxel size of 69-80 nm. 12-bit PMT detector gain was set to the highest value that included less than 10% saturated pixels in each image stack, and each frame was averaged over two scans. Stacks were projected using NIH ImageJ and subsequently false-colored, cropped, and merged using either OpenLab (Improvision) or Adobe Photoshop.

**FISH and immunofluorescence staining of embryos (Supplementary Fig. 5).** Adult transgenic animals were dissected in $dH_2O$ on charged glass slides, freeze-cracked using liquid $N_2$, and fixed at -20 °C in N,N-dimethyl formamide. Slides were then washed in PBST and slowly dehydrated to 50% formamide/2X SSCT. FISH probe diluted in hybridization buffer (10% dextran sulfate, 2 mg/ml BSA, 2X SSC 50% formamide) was added to each sample, denatured 10 min at 80ºC, and incubated overnight at 37ºC. Slides were washed three times in 50% formamide/2X SSCT at 37ºC, then in 25% formamide/2X SSCT, and finally in 2X SSCT. Slides were then rinsed in PBST and immufluorescence staining was performed as above.

**Simultaneous five-channel staining (Supplementary Fig. 1a-c).** Rabbit anti-DPY-27 and anti-SDC-2 antibodies were directly labeled with Alexa 488 and Alexa 594, respectively, using the Alexa Fluor Monoclonal Antibody Labeling Kit (Molecular Probes) following the manufacturer's instructions. The FISH probe used to mark array sequences was generated with the ULYSIS DNA labeling kit (Molecular Probes), using Alexa 660 fluorochromes to label plasmid pER723. Gravid hermaphrodites were fixed and stained as described[5] with the following changes. FISH was followed by an overnight incubation with rat anti-SDC-3. Samples were then washed in PBST and incubated 8 h with Cy3 conjugated anti-rat IgG, Alexa 488 conjugated anti-DPY-27, and Alexa 594 conjugated anti-SDC-2, and mounted in VectaShield (Vector Laboratories) with 1 μg/mL DAPI.

To empirically determine levels of fluorescence bleed-through between channels on the Leica SP-2 AOBS confocal system, samples singly stained with each fluorochrome mentioned above were subjected to spectral scanning (10 nm detector width in 2 nm steps) for every laser line used (405, 488, 546, 594, and 633 nm). The same spectral scanning was then repeated for samples stained with every combination of four fluorochromes. The resulting experimental emission spectra were evaluated using the spectral dye separation utility (included with the Leica SP-2 AOBS software), and detector settings optimized to eliminate bleed-through between channels. Each sample

was then stained with all five fluorochromes, and image stacks obtained using the optimized detector settings. Image stacks were rendered using Priism[6], and selected 3D projections were exported and subsequently false-colored using Adobe Photoshop.

**Low-copy autosomal integrants.** 0.8 μm gold beads were coated with 0.5 mg/mL plasmid containing a 4.5 kb *rex-1* insert ligated into an *unc-119* vector backbone using *Spe*I ends (pTY2293). Mixed-stage *unc-119(ed3)* hermaphrodites were bombarded with coated beads using a PDS-1000/He Biolistics Particle Delivery System (Bio-Rad) as described[3]. Independent integrations were homozygosed by cloning non-Unc hermaphrodites over multiple generations, and then subsequently determined to be autosomal or X-linked by back-crossing to wild-type males. Copy numbers of the integrated transgenes were determined by quantitative realtime PCR (Opticon 2, MJ Research) using primers amplifying both the *unc-119* part of the integration vector and the endogenous *unc-119* locus (unc-119f 5'-CGCGGCATAGAAAAAACTGG-3', unc-119r 5'- AGTATTTTCTAGGCCGTGGG-3'). Data were normalized relative to the control locus *nhr-33* (nhr-33f 5'- GTAGTTTGTGTGCGTGGGG-3', nhr-33r 5'-TTCCGGATGGACCTTGGAG-3'). Copy numbers were calculated by dividing the normalized *unc-119* level in transgenic lines by the normalized *unc-119* level in N2 animals. Two copies were subsequently subtracted from this number to account for amplification of the endogenous *unc-119* locus.

***xol-1* suppression assay.** A genetic assay to assess how strongly *rex* arrays recruit the DCC measured their abilities to rescue *xol-1* male lethality. *xol-1*(*y9*) is a null allele that causes complete XO lethality due to inappropriate binding of the DCC to the single male X and the consequent reduction of X-linked gene expression. Wild-type XO males carrying *rex*-1•33, *rex*-1•60, *rex-1*•148, or *rex-1*•241 arrays, each of which included the marker *myo-2::GFP*, were crossed into *rol-6*(*e187*) II; *xol-1*(*y9*) *unc-3*(*e151*) X hermaphrodites. Unc Grn non-Rol (XO) and non-Unc non-Rol Grn (XX) array-bearing cross progeny were scored. Percent XO *xol-1* rescue was calculated as (number *xol-1* XO array-bearing cross progeny) / (expected number of *xol-1* XO array-bearing cross progeny) X 100. The expected number of *xol-1* XO array-bearing cross progeny was calculated from data in this cross and a second cross in which XO array-bearing animals from each array strain were mated to wild-type hermaphrodites and the number of Grn male and hermaphrodite array-bearing cross progeny were counted and the fraction of XO and XX animals calculated. The expected number of *xol-1* XO progeny was then calculated by the formula: (number of non-Unc non-Rol Grn XX progeny from the first cross) (fraction of males in the second cross) / (fraction of hermahrodites in the second cross). The fraction of males in the second cross was 0.50 for *rex-1*•33, *rex-1*•60 and *rex-1*•148, and 0.65 for *rex-1*•241. The following numbers of non-Rol non-Unc Grn XX cross progeny from the set of *xol-1* crosses were the following: *rex-1*•33, N = 517; *rex-1*•60, N = 384; *rex-1*•148, N = 254; *rex-1*•241, N = 149. The following numbers of non-Rol Unc Grn X0 cross progeny from the set of *xol-1* crosses were the following: *rex-1*•33, N = 0; *rex-1*•60, N = 0; *rex-1*•148, N = 140; *rex-1*•241, N = 271. Therefore, *rex-1*•33 arrays (category 2) failed to rescue *xol-1* embryonic male lethality (0/517 expected males), as did *rex-1*•60 arrays (category 3$^-$, 0/384 expected males), but *rex-1*•148 arrays (category 3) rescued 55% of *xol-1* male lethality (140/254 expected males), and *rex-*

*1*•241 arrays (category 3) rescued 98% of *xol-1* male lethality (271/277 expected males). Most of the rescued *xol-1* males developed to L4 or adult stages.

***rex-1*•241 and *rex-1*•33 XX larval lethality.** Transgenic (green) *rex-1*•241 L4 stage hermaphrodites were transferred to fresh plates every 24 hours for 96 hours, and their entire brood scored for viability. 69% of all green eggs laid hatched and developed to adulthood, while 31% hatched but died as early (L1-L2) larvae, N = 1299 green progeny scored. The same procedure was used to assess viability of *rex-1*•33 hermaphrodites. Only 10% of *rex-1*•33 hermaphrodites died (N =1679 green progeny scored).

**Supplementary References**


1. Bailey, T. L. & Elkan, C. in Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology (AAAI Press, Menlo Park, CA, 1994).

2. Hertz, G. Z. & Stormo, G. D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15,** 563-577 (1999).

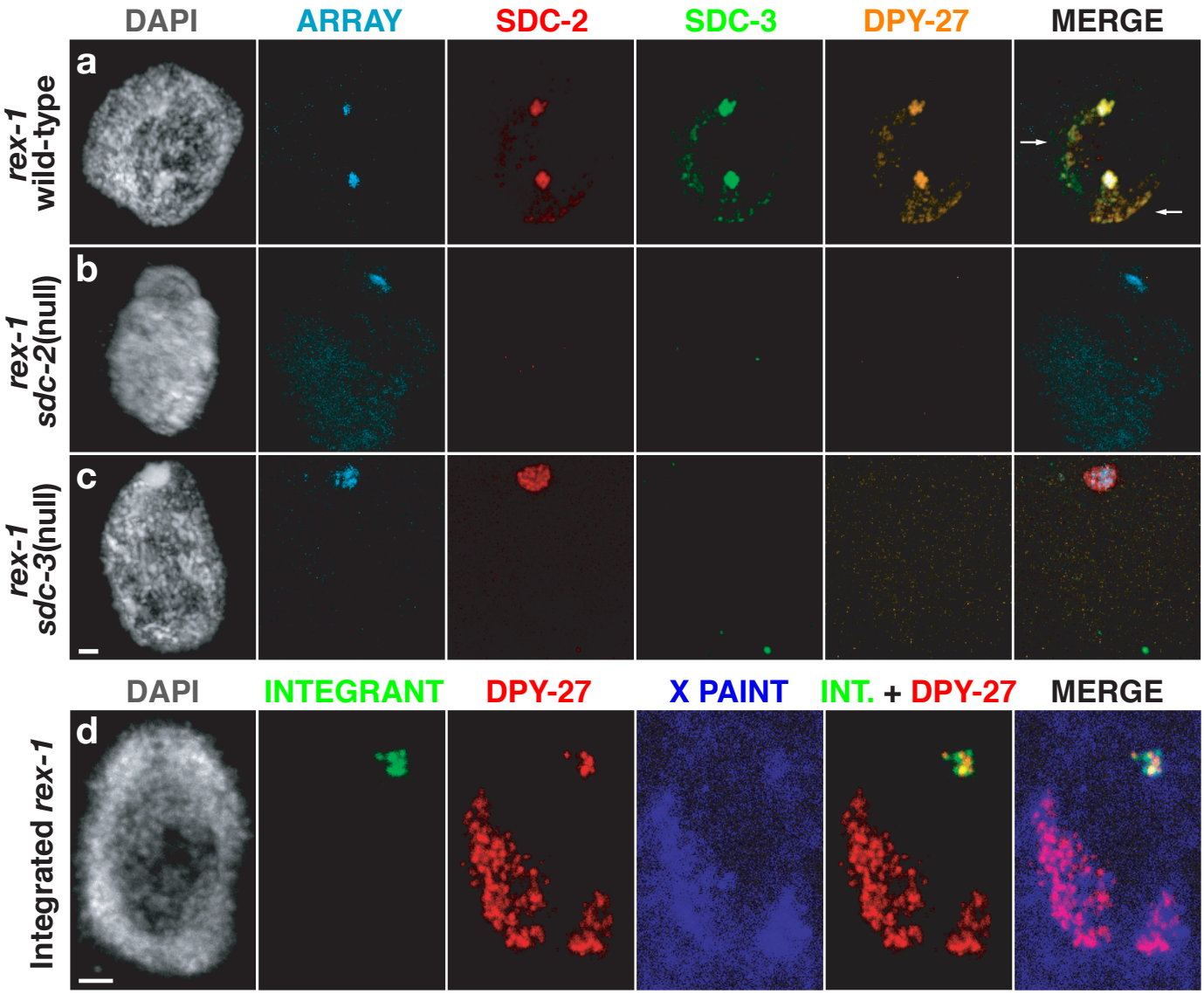
3. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157,** 1217-1226 (2001).

4. Chuang, P.-T., Albertson, D. G. & Meyer, B. J. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79,** 459-474 (1994).

5. Csankovszki, G., McDonel, P. & Meyer, B. J. Recruitment and spreading of the *C. elegans* dosage compensation complex along X chromosomes. *Science* **303,** 1182-1185 (2004).

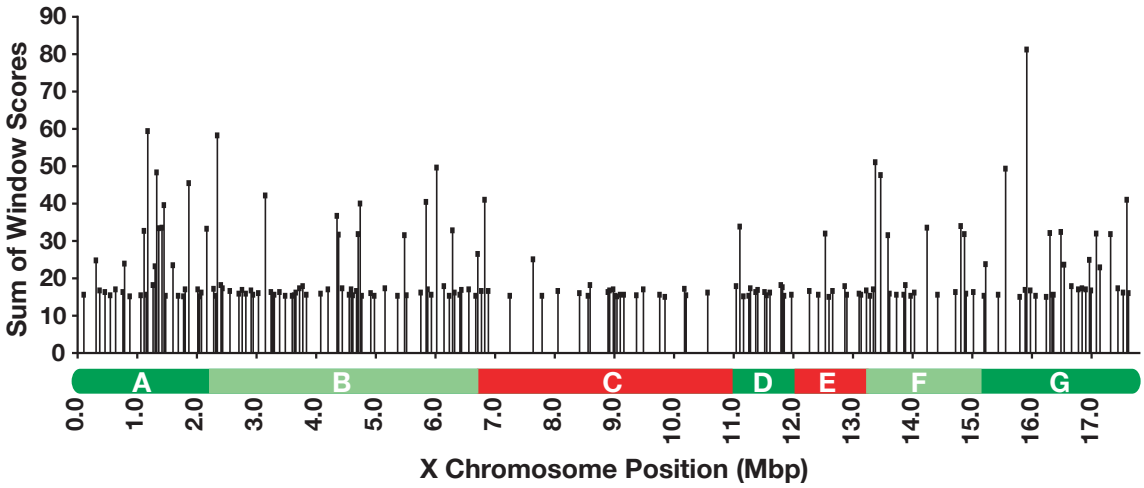
6. Chen, H., Hughes, D. D., Chan, T. A., Sedat, J. W. & Agard, D. A. IVE (Image Visualization Environment): a software platform for all three-dimensional microscopy applications. *J. Struct. Biol.* **116,** 56-60 (1996).

# Supplementary Figure 1



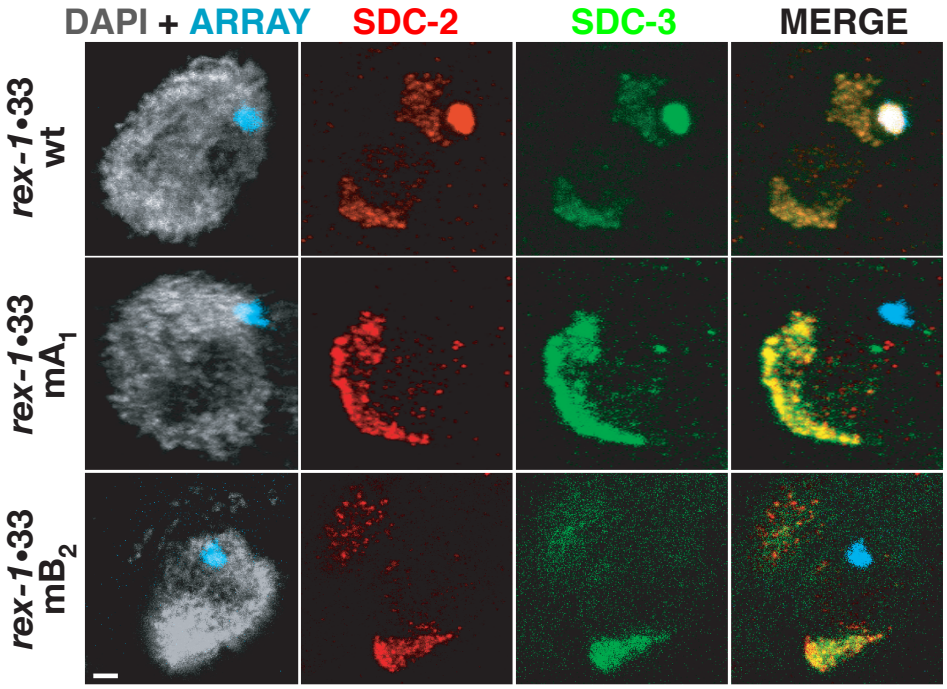

**Supplementary Figure 1 | *rex* arrays recapitulate DCC recruitment properties of X. a-c,** Confocal images of wild-type or mutant intestinal cell nuclei, all bearing *rex-1* arrays, co-stained with DAPI (grey), array FISH probe (cyan), and antibodies to dosage compensation proteins SDC-2 (red), SDC-3 (green) and DPY-27 (orange). **a,** In a wild-type nucleus, SDC-2, SDC-3, and DPY-27 co-localize with X chromosomes (white arrows) and arrays carrying 4.5 kb *rex-1* fragments. **b,** In an *sdc-2* mutant nucleus, neither *rex-1* arrays nor X chromosomes recruit the DCC. **c,** In an *sdc-3* mutant nucleus, SDC-2 binds to *rex-1* arrays, while DPY-27 does not. SDC-2 levels are known to be reduced in an *sdc-3* mutant, and that limited amount of SDC-2 is detectable on strongly recruiting *rex-1* arrays but not on X chromosomes. The same proteins direct the DCC both to *rex-1* arrays and to X chromosomes. **d,** Confocal images of an intestinal cell nucleus (grey) showing DPY-27 (red) faithfully recruited to 20 copies of *rex-1* (green) integrated at an autosomal locus by a bombardment procedure3. Each *rex-1* copy includes a 4.5 kb *rex-1* fragment ligated into the 13 kb bombardment vector. Different FISH probes detect the vector (green) and X chromosomes (blue). Scale bars, 2 μm.

# Supplementary Figure 2



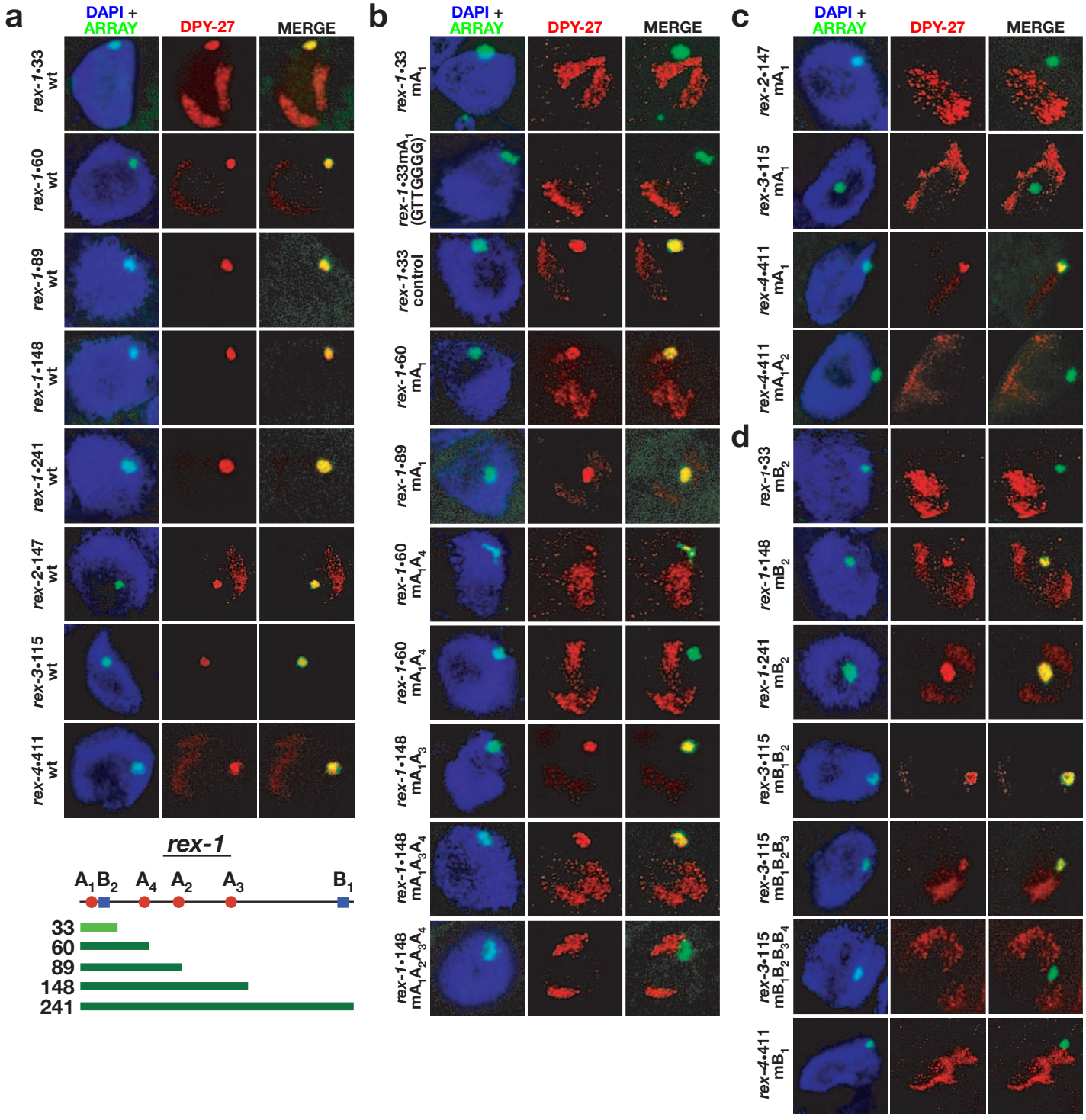

**Supplementary Figure 2 | Distribution of modeled DCC recruitment sites in 30 kb segments tiled along X.** Plot of 30 kb sequence segments along X predicted by our model to recruit the DCC. A 30 kb segment is predicted positive if it contains at least one A and one B motif, each with a raw score ≥8, within a 600 bp window (sum of clustered motif scores ≥16). The plot is juxtaposed on the recruitment map showing previously determined[5] X regions (A-G) that strongly recruit (green), weakly recruit (light green), or fail to recruit (red) the DCC when detached from X. The chi square statistical analysis of this distribution is presented in Supplementary Table 2.

# Supplementary Figure 3

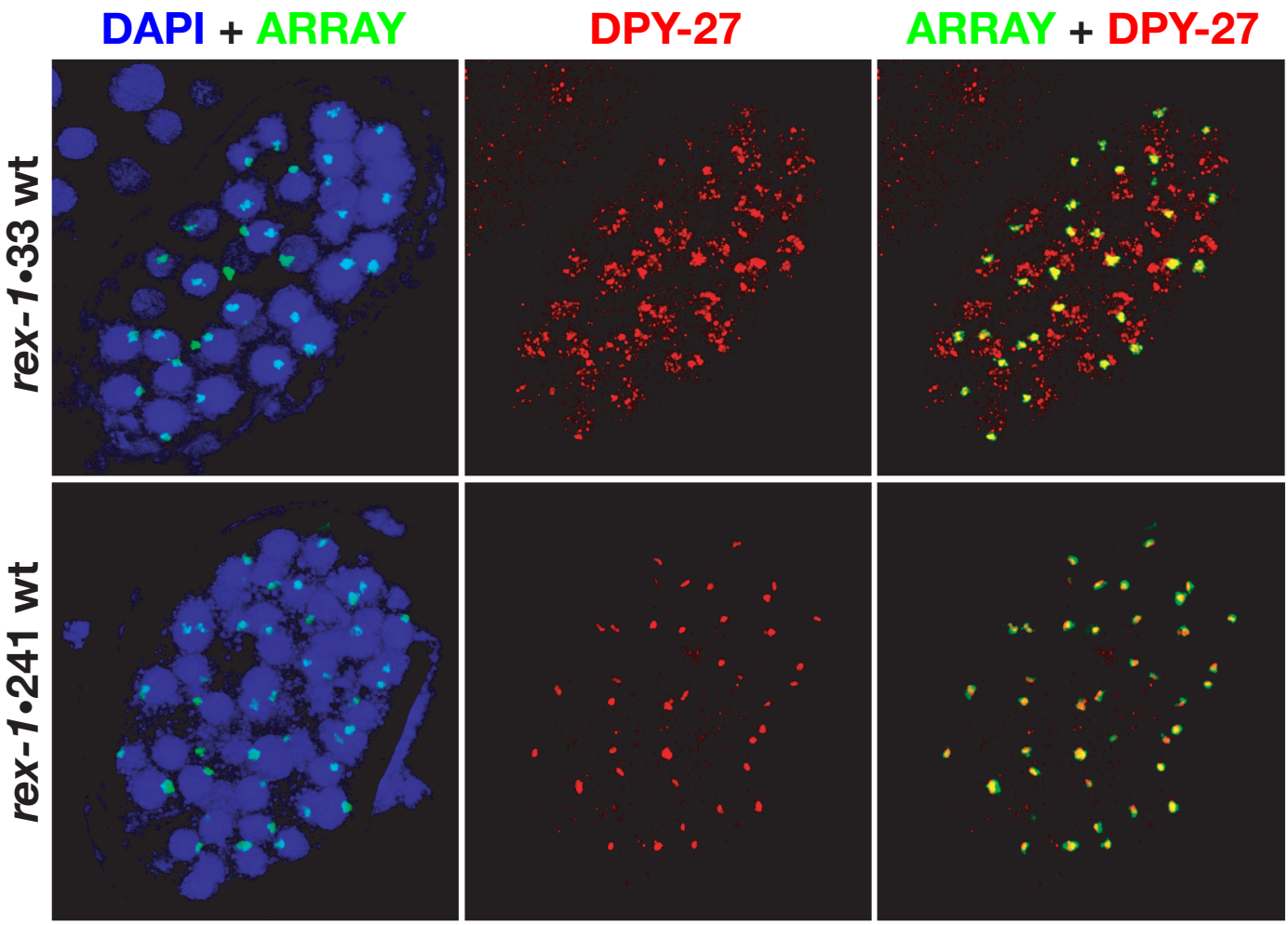



**Supplementary Figure 3 | Motifs A and B are critical for recruiting all DCC components to *rex-1*.** Intestinal cell nuclei (DAPI, grey) carrying either wild-type (*rex-1*•33) or mutant (*rex-1*•33mA1 or *rex-1*•33mB2) arrays (FISH, cyan) were co-stained with antibodies against SDC-2 (red) and SDC-3 (green), the proteins critical for targeting all DCC proteins to X. Both DCC proteins co-localized with wild-type *rex-1* arrays but neither localized to mutant *rex-1* arrays deficient in either motif A or B, revealing the importance of these motifs in directing DCC recruitment. If one of these motifs was recognized by SDC-3 alone, then mutation of that motif should not eliminate SDC-2 recruitment, since SDC-3 is not required for the binding of SDC-2 to *rex* arrays. Given that mutation of either motif A or B abolishes SDC-2 recruitment to *rex-1*•33 arrays, SDC-2 might bind one or both of these sequences directly or it might instead interact physically with another protein that specifically binds these motifs. Alternatively, motifs A and B might recruit transiently associated factors that mark *rex* sites (e.g. through histone modification) as DCC targets even before SDC-2 is expressed in the early embryo. Scale bar, 2 μm.

# Supplementary Figure 4



# Supplementary Figure 5




**Supplemental Figure 5 | *rex-1* arrays compete with the X chromosome for DCC recruitment in young embryos.** Partial confocal projections of young embryos carrying *rex-1* arrays stained with DAPI (blue), DPY-27 antibody (red), and a FISH probe (green) to identify the array. In nuclei with *rex-1*•33 arrays (category 2), DPY-27 staining is visible both on arrays and on X chromosomes. DCC recruitment to *rex-1*•241 arrays (category 3) is so strong, however, that all detectable DPY-27 is localized to arrays and absent from X. *xol-1* males carrying rex-1•33 arrays die during early development, because the arrays fail to reduce DCC binding to the male X. In contrast, *rex-1*•241 arrays compete with X so strongly for DCC binding that they rescue 99% of *xol-1* males and kill 31% of hermaphrodites.

**Supplementary Table 1 | The sequence and position (strand) of wild-type motifs in *rex* sites and the corresponding mutations**

| Motif | Strand* | Wild-Type Sequence | Mutant Sequence(s) |
|---|---|---|---|
| *rex-1* | | | |
| $A_1$ | + | GCAGGGG | GTTTTTT |
| | | | GttGGGG |
| | | | GCAttGG |
| | | | GCAGGtt |
| $A_2$ | + | GCAGGGA | GTTTTTT |
| $A_3$ | - | GCAGGGT | TTTTTTT |
| $A_4$ | + | GGAGGAG | GTTTTTT |
| $B_1$ | - | TGTAATTG | n/a |
| $B_2$ | + | GGTAATTG | GTACCAAA |
| *rex-2* | | | |
| $A_1$ | + | GCAGGGG | GTTTTTT |
| $B_1$ | - | TGTAATTG | n/a |
| $B_2$ | + | TGTAATCG | n/a |
| $B_3$ | + | GGAAAATG | n/a |
| *rex-3* | | | |
| $A_1$ | + | GCAGGGG | GTTTTTT |
| $B_1$ | + | TGTAATTG | TTTTTTTT |
| $B_2$ | + | AGCAATTG | ATTTTTTT |
| $B_3$ | + | GGCCATTG | GTTTTTTT |
| $B_4$ | - | TGTTTTTG | AAAAAAAA |
| *rex-4* | | | |
| $A_1$ | - | ACAGGGG | ATTTTTT |
| $A_2$ | - | GCGGAGG | ATTTTTT |
| $B_1$ | + | TGACATTG | TTTTTTTT |

* relative to X chromosome sequence at www.wormbase.org

**Supplementary Table S2 | A predictive model for recruitment sites corresponds with previous chromosome-wide DCC recruitment studies**

| Regions | No. of windows | Recruitment | Modeled sites | Uniformly distributed sites |
|---|---|---|---|---|
| C + E | 184 | - | 35 | 56 |
| B + F | 217 | + | 75 | 66 |
| A + D + G | 189 | ++ | 69 | 57 |

Regions of the X chromosome (A-G, shown in Supplementary Fig. 2) are sorted by their ability to recruit the DCC when detached from X. "Number of windows" refers to the total number of 30 kb windows in the pooled non-recruiting (-), weakly recruiting (+), or strongly recruiting (++) regions. "Modeled sites" refers to the number of binding sites called by our model, and "uniformly distributed sites" refers to the hypothetical number of binding sites in the pooled regions if the binding sites were distributed uniformly along the X chromosome.

Chi square analysis comparing modeled sites to uniformly distributed sites shows that modeled sites appear in significantly higher density in recruiting regions than in non-recruiting regions ($P<0.004$). Thus, the distribution of high-scoring motif A and B clusters on X corresponds well with the previous[5] DCC recruitment studies.